JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 88, 364–377 (1982)

On Some Properties of Gaussian Channels

KENJIRO YANAGI

Department of Mathematics, Yamaguchi University, Yamaguchi, Japan

Submitted by G.-C. Rota

1. INTRODUCTION

The additive Gaussian channels can be constructed in the following way. For the sake of simplicity, we consider both the input spaces and the output spaces to be a real separable Hilbert space H. Suppose that the noise source μ_0 is a Gaussian measure on H with mean vector m_0 and covariance operator R_0 and the input source μ_1 is a probability measure on H with mean vector m_1 and covariance operator R_1. Then the output source μ_2 is defined as

$$\mu_2(A) = \mu_1 \otimes \mu_0\{(x, y); x + y \in A\}, \qquad A \in \mathfrak{B},$$

where $\mu_1 \otimes \mu_0$ is the usual product meaure of μ_1 and μ_0 and $\mathfrak{B}$ is the Borel σ-field of H. Let m_2 be the mean vector of μ_2 and R_2 be the covariance operator of μ_2. Then it is easy to show that $m_2 = m_1 + m_0$ and $R_2 = R_1 + R_0$. The compound source μ_{12} derived from the input source μ_1 and the noise source μ_0 is defined by

$$\mu_{12}(B) = \mu_1 \otimes \mu_0\{(x, y);\ (x, x + y) \in B\}, \qquad B \in \mathfrak{B} \times \mathfrak{B},$$

where $\mathfrak{B} \times \mathfrak{B}$ is the Borel σ-field of $H \times H$. The capacity of additive Gaussian channels are studied in detail by Baker [7]. On the other hand, we define the complicated additive Gaussian channels in the following way. Suppose that the noise source μ_0 is a Gaussian measure on H with mean vector m_0 and covariance operator R_0 and the input source μ_1 is a probability measure on H with mean vector m_1 and covariance operator R_1. Let μ_{10} be a joint probability measure such that

$$\mu_{10}(A \times H) = \mu_1(A), \qquad A \in \mathfrak{B}$$

and

$$\mu_{10}(H \times B) = \mu_0(B), \qquad B \in \mathfrak{B}.$$

0022-247X/82/080364–14$02.00/0

The output source μ_2 is defined by

$$\mu_2(A) = \mu_{10}\{(x, y); x + y \in A\}, \qquad A \in \mathfrak{B}.$$

Let m_2 be the mean of μ_2 and R_2 be the covaraiance operator of μ_2. Then we can show that $m_2 = m_1 + m_0$ and $R_2 = R_1 + R_0 + R_{10} + R_{01}$, where R_{10} is the cross-covariance operator of μ_{10} and $R_{01} = R_{10}^*$. The compound source μ_{12} derived from the input source μ_1 and the noise source μ_0 is defined by

$$\mu_{12}(B) = \mu_{10}\{(x, y); (x, x + y) \in B\} \qquad B \in \mathfrak{B} \times \mathfrak{B}.$$

The capacity of complicated additive Gaussian channels are not yet obtained in detail.

In this paper, when we assume that μ_{10} is Gaussian in our Gaussian channels in section 3, we study the relations among the following five properties;

(a) $R_2 \geqslant R_0$ or $R_2 \leqslant R_0$,
(b) the average mutual information of μ_{10},
(c) the average mutual information of μ_{12},
(d) the strong equivalence of μ_2 and μ_0,
(e) the reproducing kernel Hilbert space of μ_0.

Also, in Section 4, we determine the maximal average mutual information under appropriate constraints.

2. Preliminaries

In this section we shall describe several useful known results relative to Gaussian measures on Hilbert spaces. Let H be a real separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and associated norm $\| \cdot \|$ and $\mathfrak{B}$ be the Borel σ-field of H. A Borel probability measure μ on $\mathfrak{B}$ that satisfies

$$\int_H \|x\|^2 \, d\mu(x) < \infty$$

defines a vector m of H and an operator R such that

$$\langle m, x \rangle = \int_H \langle y, x \rangle \, d\mu(y)$$

and

$$\langle Rx, y \rangle = \int_H \langle z - m, x \rangle \langle z - m, y \rangle \, d\mu(z).$$

The m is said to be mean vector of the measure μ. The operator R is

$$\text{linear, bounded, nonnegative, selfadjoint, and of trace-class on } H, \tag{*}$$

and we know

$$\operatorname{trace}(R) = \int_H \|x - m\|^2 \, d\mu(x).$$

In general, we call operators having the property (*) covariance operators. Let μ be a Gaussian measure on H, i.e., there exist real numbers m_x and σ_x such that

$$\mu\{y \in H: \langle x, y\rangle \leqslant a\} = \int_{-\infty}^{a} (2\pi\sigma_x)^{-1/2} \exp\{-(t - m_x)^2/(2\sigma_x)\} \, dt.$$

Then its characteristic functional $\hat{\mu}$ is given by

$$\hat{\mu}(x) = \exp\{i\langle m, x\rangle - \langle Rx, x\rangle/2\},$$

where m is the mean vector of μ, and R the covariance operator of μ. Conversely, if $m \in H$ and R is a covariance operator, then $\exp\{i\langle m, x\rangle - \langle Rx, x\rangle/2\}$ is the characteristic functional of a Gaussian measure on H. For convenience, we use the notation $\mu = [m, R]$ to denote that μ is a Gaussian measure on H with mean vector m and covariance operator R. In addition, $\mu_1 \ll \mu_2$, $\mu_1 \sim \mu_2$, and $\mu_1 \perp \mu_2$ denote that μ_1 is absolutely continuous with respect to μ_2, μ_1 and μ_2 are equivalent, and μ_1 and μ_2 are orthogonal, respectively. Also we use the notations (σc) and (τc) to denote the space of all Hilbert–Schmidt operators and the space of all trace-class operators.

Proposition 1 (Rao–Varadarajan [16]). *If* $\mu_1 = [m_1, R_1]$ *and* $\mu_2 = [m_2, R_2]$, *then* $\mu_1 \sim \mu_2$ or $\mu_1 \perp \mu_2$. Also, $\mu_1 \sim \mu_2$ if and only if

(a) $m_2 - m_1 \in \operatorname{range}(R_1^{1/2}) = \operatorname{range}(R_2^{1/2})$ *and*

(b) $R_2 = R_1^{1/2}(I + T)R_1^{1/2}$ *for certain* $T \in (\sigma c)$.

Proposition 2. *Let* $\mu_1 = [m_1, R_1]$ *and* $\mu_2 = [m_2, R_2]$. *If* $\mu_1 \sim \mu_2$, *then*

$$\frac{d\mu_2}{d\mu_1}(x) = \exp\left\{\frac{1}{2}\sum_{k,j} \langle T(I + T)^{-1} e_k, e_j\rangle \left[\frac{\langle x - m_2, e_k\rangle\langle x - m_2, e_j\rangle}{(\lambda_k \lambda_j)^{1/2}} - \delta_{kj}\right]\right.$$
$$\left. + \frac{1}{2}\sum_k \left[\frac{\tau_k}{1 + \tau_k} - \log(1 + \tau_k)\right]\right.$$

$$+\sum_k \lambda_k^{-1}\langle x-m_1, e_k\rangle\langle m_2-m_1, e_k\rangle$$
$$-\frac{1}{2}\sum_k \lambda_k^{-1}\langle m_2-m_1, e_k\rangle^2\Bigg\},$$

where $\{\lambda_k\}$ are nonzero eigenvalues of R_1, $\{e_k\}$ are corresponding orthonormal eigenvectors of R_1, and $\{\tau_k\}$ are nonzero eigenvalues of T. Also we obtain

$$\int_H \log\frac{d\mu_2}{d\mu_1}(x)\,d\mu_2(x)$$
$$=\frac{1}{2}\sum_n \{\tau_n-\log(1+\tau_n)\}+\frac{1}{2}\sum_n \lambda_n^{-1}\langle m_2-m_1, e_n\rangle^2.$$

Remark 1. In Propositions 1 and 2, if we set T to be zero on null(R_1), then T is uniquely determined.

When $T\in(\tau c)$ in Proposition 1, we use the notation $\mu_1\sim^s\mu_2$ to denote that μ_1 and μ_2 are strongly equivalent. Suppose that H_1, H_2 are real separable Hilbert spaces with inner products $\langle\cdot,\rangle_1$, $\langle\cdot,\cdot\rangle_2$; associated norms $\|\cdot\|_1$, $\|\cdot\|_2$; and $\mathfrak{B}_1=\mathfrak{B}(H_1)$, $\mathfrak{B}_2=\mathfrak{B}(H_2)$ the Borel σ-fields of H_1, H_2, respectively. Denote $H_1\times H_2$ the real separable Hilbert space under the inner product $[(u,v),\ (y,z)]=\langle u,y\rangle_1+\langle v,z\rangle_2$ and associated norm $|||(x,y)|||^2=[(x,y),\ (x,y)]$. Moreover, the norm-open sets obtained via this inner product generate the Borel σ-field $\mathfrak{B}_1\times\mathfrak{B}_2=\mathfrak{B}(H_1\times H_2)$. Let μ_1,μ_2 be Borel probability measures on $\mathfrak{B}_1$, $\mathfrak{B}_2$ and μ_{12} be a joint probability measure on $\mathfrak{B}_1\times\mathfrak{B}_2$ such that μ_{12} has μ_1,μ_2 as projections on H_1, H_2, respectively; i.e., $\mu_1(A)=\mu_{12}(A\times H_2)$ and $\mu_2(B)=\mu_{12}(H_1\times B)$. If $\int_{H_1}\|x\|_1^2\,d\mu_1(x)<\infty$ and $\int_{H_2}\|x\|_2^2\,d\mu_2(x)<\infty$, then we can define a unique cross-covariance operator R_{12}: $H_2\to H_1$ by

$$\langle R_{12}y,x\rangle_1=\int_{H_1\times H_2}\langle u-m_1,x\rangle_1\,\langle v-m_2,y\rangle_2\,d\mu_{12}(u,v),$$

where m_1 and m_2 denote the mean vectors of μ_1 and μ_2, respectively.

PROPOSITION 3 (Baker [4]). *$R_{12}=R_1^{1/2}VR_2^{1/2}$, where V: $H_2\to H_1$ is a bounded linear operator such that $\|V\|\leqslant 1$. If we set V to satisfy the condition* null$(R_2)\subset$ null(V) *and* $\overline{\text{range}(V)}\subset\overline{\text{range}(R_1)}$, *then V is uniquely determined.*

When $\mu_1\otimes\mu_2$ is the usual product measure on $\mathfrak{B}_1\times\mathfrak{B}_2$ of μ_1 and μ_2, the average mutual information $I(\mu_{12})$ of the measure μ_{12} with respect to $\mu_1 \otimes \mu_2$ is defined as follows: If $\mu_{12} \ll \mu_1 \otimes \mu_2$,

$$I(\mu_{12}) = \int_{H_1 \times H_2} \log \frac{d\mu_{12}}{d\mu_1 \otimes \mu_2}(x, y)\, d\mu_{12}(x, y),$$

and otherwise, $I(\mu_{12}) = \infty$.

PROPOSITION 4 (Baker [4]). *Suppose that μ_{12} is Gaussian. Then the following are equivalent*:

(a) $\mu_{12} \sim \mu_1 \otimes \mu_2$;

(b) $V \in (\sigma c)$ and $\|V\| < 1$;

(c) $I(\mu_{12}) < \infty$.

The following result will be used in this and succeeding sections.

PROPOSITION 5 (Douglas [11]). *Suppose that R_1 and R_2 are covariance operators. Then*

(5A) $\mathrm{range}(R_1^{1/2}) \subset \mathrm{range}(R_2^{1/2})$ *if and only if there exists a bounded linear operator C such that $R_1^{1/2} = R_2^{1/2}C$*;

(5B) $\mathrm{range}(R_1^{1/2}) = \mathrm{range}(R_2^{1/2})$ *if and only if there exists a bounded linear operator C having bounded inverse such that $R_1^{1/2} = R_2^{1/2}C$*;

(5C) $R_1^{1/2} = R_2^{1/2}C$, *where $\|C\| \leqslant 1$, if and only if $R_1 \leqslant R_2$.*

Remark 2. In Proposition 5, if we set C to satisfy the condition $\overline{\mathrm{range}(C)} \subset \overline{\mathrm{range}(R_2)}$, then C is uniquely determined.

PROPOSITION 6 (Baker [5]). *Let μ be a non-Gaussian measure with mean vector m and covariance operator R and let $T: H \to H$ be a bounded linear operator. We consider the conditions*:

(a) $\mu[\mathrm{range}(T)] = 1$;

(b) $m \in \mathrm{range}(T)$ and $R = TST^*$, where $S \in (\tau c)$;

(c) $m \in \mathrm{range}(T)$ and $R^{1/2} = TU$, where $U \in (\sigma c)$.

Then the following implications hold: (a) $\Rightarrow$ (b) $\Leftrightarrow$ (c).

In particular when μ is Gaussian, (a) $\Leftrightarrow$ (b) $\Leftrightarrow$ (c). If we set S to satisfy the condition $\overline{\mathrm{range}(S)} \subset \overline{\mathrm{range}(T)}$, then S is a covariance operator determined uniquely. Also if we set U to satisfy the condition $\overline{\mathrm{range}(U)} \subset \overline{\mathrm{range}(T)}$, then U is uniquely determined.

3. Some Relations

In this section, we assume that μ_{10} is Gaussian in the complicated Gaussian channels defined in Section 1. Then the measures μ_1, μ_0, μ_2 and μ_{12} are also Gaussian. And we can assume that $m_1 = m_0 = 0$. By Proposition 3, we obtain $R_{10} = R_1^{1/2} V R_0^{1/2}$, where $\|V\| \leqslant 1$, $\text{null}(R_0) \subset \text{null}(V)$, and $\overline{\text{range}(V)} \subset \overline{\text{range}(R_1)}$. Similarly, $R_{12} = R_1^{1/2} U R_2^{1/2}$, where $\|U\| \leqslant 1$, $\text{null}(R_2) \subset \text{null}(U)$, and $\overline{\text{range}(U)} \subset \overline{\text{range}(R_1)}$. The following lemma is useful.

Lemma 1. $R_2^{1/2}(I - U^*U)R_2^{1/2} = R_0^{1/2}(I - V^*V)R_0^{1/2}$.

Proof. Since $m_2 = m_1 + m_0 = 0$, the following equalities hold:

$$\begin{aligned}
\langle R_{12} v, u\rangle &= \int_{H\times H} \langle x, u\rangle\langle y, v\rangle \, d\mu_{12}(x, y) \\
&= \int_{H\times H} \langle p, u\rangle\langle p+q, v\rangle \, d\mu_{10}(p, q) \\
&= \int_{H\times H} \langle p, u\rangle\langle p, v\rangle \, d\mu_{10}(p, q) \\
&\quad + \int_{H\times H} \langle p, u\rangle\langle q, v\rangle \, d\mu_{10}(p, q) \\
&= \int_{H} \langle p, u\rangle\langle p, v\rangle \, d\mu_1(p) \\
&\quad + \int_{H\times H} \langle p, u\rangle\langle q, v\rangle \, d\mu_{10}(p, q) \\
&= \langle R_1 v, u\rangle + \langle R_{10} v, u\rangle \\
&= \langle (R_1 + R_{10}) v, u\rangle.
\end{aligned}$$

Hence $R_{12} = R_1 + R_{10}$. Thus $R_1 + R_{10} = R_1^{1/2} U R_2^{1/2}$ and so $R_1^{1/2}(R_1^{1/2} + VR_0^{1/2} - UR_2^{1/2}) = 0$. Then $(R_1^{1/2} + VR_0^{1/2} - UR_2^{1/2})x \in \text{null}(R_1) = \overline{\text{range}(R_1)}^{\perp}$ for every $x \in H$. But since we assume that $\overline{\text{range}(V)} \subset \overline{\text{range}(R_1)}$ and $\overline{\text{range}(U)} \subset \overline{\text{range}(R_1)}$, $(R_1^{1/2} + VR_0^{1/2} - UR_2^{1/2})x \in \overline{\text{range}(R_1)}$ for every $x \in H$. Hence $(R_1^{1/2} + VR_0^{1/2} - UR_2^{1/2})x = 0$ for every $x \in H$. Thus $R_1^{1/2} + VR_0^{1/2} - UR_2^{1/2} = 0$. We obtain $R_2^{1/2} U^* U R_2^{1/2} = (R_1^{1/2} + VR_0^{1/2})^* (R_1^{1/2} + VR_0^{1/2}) = R_1 + R_{10} + R_{01} + R_0^{1/2} V^* V R_0^{1/2}$. Since $R_2 = R_1 + R_0 + R_{10} + R_{01}$, $R_2 - R_0 + R_0^{1/2} V^* V R_0^{1/2} = R_2^{1/2} U^* U R_2^{1/2}$ and we obtain the equality $R_2^{1/2}(I - U^*U)R_2^{1/2} = R_0^{1/2}(I - V^*V)R_0^{1/2}$. Q.E.D.

Now we shall obtain the following

THEOREM 1. *Suppose that $I(\mu_{10}) < \infty$. We consider the conditions*:

(A) $R_2 \leqslant R_0$;

(B) $\mu_2 \sim^s \mu_0$;

(C) $I(\mu_{12}) < \infty$;

(D) $\mu_1[\text{range}(R_0^{1/2})] = 1$.

Then the following implications hold: (A) $\Rightarrow$ (B) $\Leftrightarrow$ (C) $\Leftrightarrow$ (D).

Remark 3. The equivalence of (B)–(D) have been proved by Baker [9]. We shall give an another simple proof of those. $I(\mu_{10}) < \infty$ and $I(\mu_{12}) < \infty$ do not necessarily imply $R_2 \leqslant R_0$. As an example we may take the operators U, $V \in (\sigma c)$ such that $\|U\| < 1$, $\|V\| < 1$, and $\|(I - V^*V)^{1/2} T(I - U^*U)^{-1/2}\| > 1$ for some operator T.

Proof of Theorem 1. We show that (A) $\Rightarrow$ (B) $\Rightarrow$ (C) $\Rightarrow$ (D) $\Rightarrow$ (B).

(A) $\Rightarrow$ (B): By Proposition 5, $R_2 \leqslant R_0$ implies $\text{range}(R_2^{1/2}) \subset \text{range}(R_0^{1/2})$. Since $I(\mu_{10}) < \infty$, we have $V \in (\sigma c)$, $\|V\| < 1$ by Proposition 4. By Lemma 1 and Proposition 5, we obtain $\text{range}(R_0^{1/2}) \subset \text{range}(R_2^{1/2})$. Then $\text{range}(R_2^{1/2}) = \text{range}(R_0^{1/2})$. Lemma 1 and $R_2 \leqslant R_0$ imply $R_0^{1/2}V^*VR_0^{1/2} \geqslant R_2^{1/2}U^*UR_2^{1/2}$. When we set $R_0^{1/2} = R_2^{1/2}B$, then $R_2^{1/2}BV^*VB^*R_2^{1/2} \geqslant R_2^{1/2}U^*UR_2^{1/2}$ and $\langle (BV^*VB^* - U^*U)R_2^{1/2}x, R_2^{1/2}x\rangle \geqslant 0$. Since B^* and U are 0 on $\text{null}(R_2)$, $BV^*VB^* \geqslant U^*U$. Then $U \in (\sigma c)$ by $V \in (\sigma c)$. Thus $R_0 = R_2^{1/2}(I + BV^*VB^* - U^*U)R_2^{1/2}$, where $BV^*VB^* - U^*U \in (\tau c)$ and $\text{rnage}(R_2^{1/2}) = \text{range}(R_0^{1/2})$. Hence $\mu_2 \sim^s \mu_0$.

(B) $\Rightarrow$ (C): In proof of (A) $\Rightarrow$ (B), $\text{range}(R_0^{1/2}) \subset \text{range}(R_2^{1/2})$. By Proposition 5, there exists a bounded linear operator B such that $R_0^{1/2} = R_2^{1/2}B$. Since $R_0^{1/2}(I - V^*V)R_0^{1/2} = R_2^{1/2}(I - U^*U)R_2^{1/2}$ by Lemma 1, we obtain the following equalities:

$$\begin{aligned} R_0 &= R_0^{1/2}V^*VR_0^{1/2} + R_2^{1/2}(I - U^*U)R_2^{1/2} \\ &= R_2^{1/2}BV^*VB^*R_2^{1/2} + R_2^{1/2}(I - U^*U)R_2^{1/2} \\ &= R_2^{1/2}(I + BV^*VB^* - U^*U)R_2^{1/2}. \end{aligned}$$

Then $\mu_0 \sim^s \mu_2$ implies

$$BV^*VB^* - U^*U \in (\tau c) \tag{1}$$

$$\text{range}(R_0^{1/2}) = \text{range}(R_2^{1/2}). \tag{2}$$

By (1), $U^*U \in (\tau c)$ and so $U \in (\sigma c)$. By (2) and Proposition 5, $I - U^*U$ has a bounded inverse. Since $\|U\|^2 = \|U^*U\| < 1$, $I(\mu_{12}) < \infty$ by Proposition 4.

(C) $\Rightarrow$ (D): By Lemma 1, $R_2^{1/2}(I - U^*U)R_2^{1/2} = R_0^{1/2}(I - V^*V)R_0^{1/2}$. Since $I(\mu_{10}) < \infty$ and $I(\mu_{12}) < \infty$, we have V, $U \in (\sigma c)$ such that $\|V\| < 1$,

$\|U\| < 1$ by Proposition 4. Proposition 5 implies range$(R_2^{1/2}) =$ range$(R_0^{1/2})$. Then there exists a bounded linear operator B having bounded inverse with $R_2^{1/2} = R_0^{1/2}B$. Since $R_1^{1/2} = UR_2^{1/2} - VR_0^{1/2}$, $R_1^{1/2} = (R_1^{1/2})^* = R_2^{1/2}U^* - R_0^{1/2}V^* = R_0^{1/2}BU^* - R_0^{1/2}V^* = R_0^{1/2}(BU^* - V^*)$, where $BU^* - V^* \in (\sigma c)$ by U, $V \in (\sigma c)$. Hence $\mu_1[\text{range}(R_0^{1/2})] = 1$ by Proposition 6.

(D) $\Rightarrow$ (B): Suppose that $\mu_1[\text{range}(R_0^{1/2})] = 1$. By Proposition 6, $R_1^{1/2} = R_0^{1/2}$, where $P \in (\sigma c)$. Then $R_2 = R_1 + R_0 + R_{10} + R_{01} = R_0^{1/2}PP^*R_0^{1/2} + R_0 + R_0^{1/2}PVR_0^{1/2} + R_0^{1/2}V^*P^*R_0^{1/2} = R_0^{1/2}(I + PV + V^*P^* + PP^*)R_0^{1/2}$. Since P, $V \in (\sigma c)$, $PV + V^*P^* + PP^* \in (\tau c)$. By Proposition 5, range$(R_2^{1/2}) \subset$ range$(R_0^{1/2})$. On the other hand, range$(R_0^{1/2}) \subset$ range$(R_2^{1/2})$ by Lemma 1 and Proposition 5. Then range$(R_2^{1/2}) =$ range$(R_0^{1/2})$. We obtain $\mu_2 \sim^s \mu_0$. Q.E.D.

Moreover we have the following

THEOREM 2. *Suppose that* $I(\mu_{12}) < \infty$. *We consider the conditions*:

(A) $R_2 \geqslant R_0$;

(B) $\mu_2 \sim^s \mu_0$;

(C) $I(\mu_{10}) < \infty$;

(D) $\mu_1[\text{range}(R_0^{1/2})] = 1$.

Then the following implications hold: (A) $\Rightarrow$ (B) $\Leftrightarrow$ (C) $\Rightarrow$ (D).

Remark 4. Baker [9] proved only that (A) $\Rightarrow$ (B)–(D). $I(\mu_{10}) < \infty$ and $I(\mu_{12}) < \infty$ do not necessarily imply $R_2 \geqslant R_0$. As an example we may take the operators U, $V \in (\sigma c)$ such that $\|U\| < 1$, $\|V\| < 1$, and $\|(I - U^*U)^{1/2} T(I - V^*V)^{-1/2}\| > 1$ for some operator T. $I(\mu_{12}) < \infty$ and $\mu_1[\text{range}(R_0^{1/2})] = 1$ do not necessarily imply $I(\mu_{10}) < \infty$. As an example we may take the operators U, $V \in (\sigma c)$ such that $\|U\| < 1$, $\|V\| = 1$, and $R_1^{1/2} = R_0^{1/2}V$.

Proof of Theorem 2. By Theorem 1, it is clear that (C) $\Rightarrow$ (B) and (C) $\Rightarrow$ (D). We may show that (A) $\Rightarrow$ (B) $\Rightarrow$ (C). (A) $\Rightarrow$ (B): By Proposition 5, $R_2 \geqslant R_0$ implies range$(R_0^{1/2}) \subset$ range$(R_2^{1/2})$. Since $I(\mu_{12}) < \infty$, we have $U \in (\sigma c)$, $\|U\| < 1$ by Proposition 4. By Lemma 1 and Proposition 5, we obtain range$(R_2^{1/2}) \subset$ range$(R_0^{1/2})$. Then range$(R_2^{1/2}) =$ range$(R_0^{1/2})$. Lemma 1 and $R_2 \geqslant R_0$ imply $R_2^{1/2}U^*UR_2^{1/2} \geqslant R_0^{1/2}V^*VR_0^{1/2}$. When we set $R_2^{1/2} = R_0^{1/2}T$, then $R_0^{1/2}\ TU^*UT^*R_0^{1/2} \geqslant R_0^{1/2}V^*VR_0^{1/2}$ and so $\langle (TU^*UT^* - V^*V)R_0^{1/2}x,\ R_0^{1/2}x \rangle \geqslant 0$. Since T^* and V are 0 on null(R_0), $TU^*UT^* \geqslant V^*V$. Then $V \in (\sigma c)$ by $U \in (\sigma c)$. Thus $R_2 = R_0^{1/2}(I + TU^*UT^* - V^*V)R_0^{1/2}$, where $TU^*UT^* - V^*V \in (\tau c)$ and range$(R_2^{1/2}) =$ range$(R_0^{1/2})$. Hence $\mu_2 \sim^s \mu_0$.

(B) $\Rightarrow$ (C): In proof of (A) $\Rightarrow$ (B), $\mathrm{range}(R_2^{1/2}) \subset \mathrm{range}(R_0^{1/2})$. By Proposition 5, there exists a bounded linear operator B such that $R_2^{1/2} = R_0^{1/2}B$. Since $R_2^{1/2}(I - U^*U)R_2^{1/2} = R_0^{1/2}(I - U^*U)R_0^{1/2}$ by Lemma 1, we obtain the following equalities:

$$\begin{aligned} R_2 &= R_2^{1/2}U^*UR_2^{1/2} + R_0^{1/2}(I - V^*V)R_0^{1/2} \\ &= R_0^{1/2}BU^*UB^*R_0^{1/2} + R_0^{1/2}(I - V^*V)R_0^{1/2} \\ &= R_0^{1/2}(I + BU^*UB^* - V^*V)R_0^{1/2}. \end{aligned}$$

Then $\mu_2 \sim^s \mu_0$ implies

$$BU^*UB^* - V^*V \in (\tau c) \tag{3}$$

$$\mathrm{range}(R_2^{1/2}) = \mathrm{range}(R_0^{1/2}). \tag{4}$$

By (3), $V^*V \in (\tau c)$ and so $V \in (\sigma c)$. By (4) and Proposition 5, $I - V^*V$ has a bounded inverse. Since $\|V\|^2 = \|V^*V\| < 1$, $I(\mu_{10}) < \infty$ by Proposition 4.

Q.E.D.

Remark 5. In [9], Baker proved that in order that either $I(\mu_{12}) < \infty$ or $\mu_2 \sim^s \mu_0$, it is necessary that $\mathrm{range}(R_1^{1/2}) \subset \mathrm{range}(R_0^{1/2})$. In this proposition the condition $\mu_2 \sim^s \mu_0$ can be changed by weaker condition $\mu_2 \sim \mu_0$.

Remark 6. We can easily show the following properties:

(a) If $\mu_1[\mathrm{range}(R_0^{1/2})] = 1$, then $R_2 \geqslant R_0$ implies $\mu_2 \sim \mu_0$.

(b) If $\mu_1[\mathrm{range}(R_0^{1/2})] = 1$, then $V \in (\sigma c)$ is equivalent to $U \in (\sigma c)$.

(c) If $\mu_1[\mathrm{range}(R_0^{1/2})] = 1$, then $I(\mu_{10}) < \infty$ implies $I(\mu_{12}) < \infty$ and $\mu_2 \sim^s \mu_0$.

4. Maximal Average Mutual Information

In this section, we require the same constraints as previous sections, and we consider appropriate constraints and determine the maximal average mutual information under them. Suppose that $\dim[H] \geqslant M$, then we can induce the constraints for R_1 and R_{10} in the following:

(a) $R_1^{1/2} = R_0^{1/2}S$ for certain $S \in (\sigma c)$,

(b) $R_{10} = R_1^{1/2}VR_0^{1/2}$ for certain $V \in (\sigma c)$ such that $\|V\| < 1$, where S and V satisfy

(c) $\sigma[(I - V^*V)^{-1/2}(S + V^*)] \leqslant P^{1/2} < \infty$, where $\sigma[\cdot]$ is the Hilbert–Schmidt norm,

(d) $(S + V^*)(S^* + V)$ and V^*V commute and $\dim[\mathrm{range}(V^*V)] \leqslant M$.

If μ_{10} is Gaussian, then μ_{10} is uniquely determined by R_1 and R_{10}. Since μ_{12} is defined by μ_{10}, we shall use the notations $Q = \{\mu_{12}, R_1$ and R_{10} satisfy (a)–(d)$\}$ and $C(Q) = \sup\{I(\mu_{12}); \mu_{12} \in Q\}$ to obtain theorems on maximal average mutual information under constraints Q.

Remark 7. For simplicity, we can and may assume that $\overline{\text{range}(R_0)} = H$. Then S is uniquely determined. Constraint (b) assures that $\mu_{12} \sim \mu_1 \otimes \mu_2$ and constraint (c) assures that $C(Q)$ is finite. Constraint (d) is necessary to calculate $C(Q)$.

The following lemma is useful.

LEMMA 2. *Under constraints Q, $I(\mu_{12}) = -\frac{1}{2}\sum_n \log(1-\delta_n)$, where $\{\delta_n\}$ are nonzero eigenvalues of U^*U.*

Proof. Let $\mathfrak{R}_{12}, \mathfrak{R}_{1\otimes 2}$ be the covariance operators of $\mu_{12}, \mu_1 \otimes \mu_2$, respectively. Let $\mathfrak{U}$: $H \times H \to H \times H$ be the selfadjoint operator defined by $\mathfrak{U}(u, v) = (Uv, U^*u)$, and let $\mathfrak{I}$ be the identity in $H \times H$. The $\mathfrak{U}$ is a selfadjoint linear operator with $\|\mathfrak{U}\| \leqslant 1$. Because μ_{10} is Gaussian, μ_{12} and $\mu_1 \otimes \mu_2$ are also Gaussian. By constraint (b), $I(\mu_{10}) < \infty$. Using Theorem 1, $I(\mu_{12}) < \infty$. Then we can obtain that $\mathfrak{R}_{12} = \mathfrak{R}_{1\otimes 2}^{1/2}(\mathfrak{I} + \mathfrak{U})\mathfrak{R}_{1\otimes 2}^{1/2}$, where $\mathfrak{U} \in (\sigma c)$ and $\mathfrak{I} + \mathfrak{U}$ is invertible. By Proposition 2, $I(\mu_{12}) = \frac{1}{2}\sum_n \{\tau_n - \log(1+\tau_n)\}$, where $\{\tau_n\}$ are nonzero eigenvalues of $\mathfrak{U}$. Using the fact that $\{\tau_n\}$ are eigenvalues of $\mathfrak{U}$ of multiplicity N, if and only if $\{-\tau_n\}$ are eigenvalues of $\mathfrak{U}$ of multiplicity N, we obtain

$$I(\mu_{12}) = \frac{1}{2}\sum_n \{\tau_n - \log(1+\tau_n)\} = -\frac{1}{2}\sum_n \log(1-\tau_n^2),$$

where $\{\tau_n^2\}$ are nonzero eigenvalues of U^*U. Q.E.D.

We can now obtain the following theorem on the maximal average mutual information.

THEOREM 3. *If $M < \infty$, then $C(Q) = (M/2)\log(1 + (P/M))$ and the maximum is attained.*

Proof. By constraints (a) and (b),

$$\begin{aligned} R_2 &= R_1 + R_0 + R_{10} + R_{01} \\ &= R_0^{1/2}SS^*R_0^{1/2} + R_0 + R_1^{1/2}VR_0^{1/2} + R_0^{1/2}V^*R_1^{1/2} \\ &= R_0^{1/2}SS^*R_0^{1/2} + R_0 + R_0^{1/2}SVR_0^{1/2} + R_0^{1/2}V^*S^*R_0^{1/2} \\ &= R_0^{1/2}(I + SS^* + SV + V^*S^*)R_0^{1/2}. \end{aligned}$$

By Theorem 1, μ_2 and μ_0 are strongly equivalent. Then $\overline{\text{range}(R_2^{1/2})} = \overline{\text{range}(R_0^{1/2})}$. Hence $R_2^{1/2} = R_0^{1/2}\ (I + SS^* + SV + V^*S^*)^{1/2} W^*$, where W is a partial isometry such that W is isometric on $\overline{\text{range}(R_2)} = \overline{\text{range}(R_0)}$ and zero on $\text{null}(R_2)$. We can obtain $R_2^{1/2} U^*U R_2^{1/2} = R_0^{1/2}(I + SS^* + SV + V^*S^*)^{1/2}\ W^*U^*UW(I + SS^* + SV + V^*S^*)^{1/2} R_0^{1/2}$. On the other hand, by Lemma 1,

$$\begin{aligned} R_2^{1/2} U^*U R_2^{1/2} &= R_2 - R_0^{1/2}(I - V^*V) R_0^{1/2} \\ &= R_0^{1/2}(I + SS^* + SV + V^*S^*) R_0^{1/2} - R_0^{1/2}(I - V^*V) R_0^{1/2} \\ &= R_0^{1/2}(SS^* + SV + V^*S^* + V^*V) R_0^{1/2}. \end{aligned}$$

Since we can assume that $\overline{\text{range}(R_0)} = H$, we obtain

$$\begin{aligned} &(I + SS^* + SV + V^*S^*)^{1/2} W^*U^*UW(I + SS^* + SV + V^*S^*)^{1/2} \\ &\quad = (S + V^*)(S^* + V), \end{aligned}$$

and also since we can assume that W is a unitary operator on H,

$$\begin{aligned} U^*U = {} & W(I + SS^* + SV + V^*S^*)^{-1/2} \\ & \times (S + V^*)(S^* + V)(I + SS^* + SV + V^*S^*)^{-1/2} W^*. \end{aligned}$$

Then U^*U has the same point spectrum as

$$(I + SS^* + SV + V^*S^*)^{-1/2}(S + V^*)(S^* + V)(I + SS^* + SV + V^*S^*)^{-1/2}.$$

Let $\{e_n\}$ be the eigenvectors of $(S + V^*)(S^* + V)$ and $\{\lambda_n\}$ the corresponding eigenvalues of $(S + V^*)(S^* + V)$. By constraint (d), V^*V has the eigenvectors $\{e_n\}$ and corresponding eigenvalues $\{\mu_n\}$. The U^*U has the eigenvalues $\{\lambda_n/(1 + \lambda_n - \mu_n)\}$. We have by Lemma 2,

$$I(\mu_{12}) = -\frac{1}{2}\sum_n \log\left(1 - \frac{\lambda_n}{1 + \lambda_n - \mu_n}\right) = \frac{1}{2}\sum_n \log\left(1 + \frac{\lambda_n}{1 - \mu_n}\right).$$

And by constraint (d), $\{\lambda_n/(1 - \mu_n)\}$ are the eigenvalues of

$$(I - V^*V)^{-1/2}(S + V^*)(S^* + V)(I - V^*V)^{-1/2}.$$

Let $\lambda_1/(1 - \mu_1), \lambda_2/(1 - \mu_2), \ldots, \lambda_K/(1 - \mu_K)$ be nonzero eigenvalues of

$$(I - V^*V)^{-1/2}(S + V^*)(S^* + V)(I - V^*V)^{-1/2}.$$

We have that

$$\begin{aligned}
&\sigma[(I-V^*V)^{-1/2}(S+V^*)]^2 \\
&\quad =\sigma[(S^*+V)(I-V^*V)^{-1/2}]^2 \\
&\quad =\mathrm{trace}[(I-V^*V)^{-1/2}(S+V^*)(S^*+V)(I-V^*V)^{-1/2}] \\
&\quad =\sum_{n=1}^{K}\frac{\lambda_n}{1-\mu_n}=P_1\leqslant P.
\end{aligned}$$

Then it is sufficient to obtain the maximum of $\frac{1}{2}\sum_{n=1}^{K}\log(1+(\lambda_n/(1-\mu_n)))$ when $\sum_{n=1}^{K}\lambda_n/(1-\mu_n)\leqslant P$. We wish to choose K, P_1, and $\lambda_1/(1-\mu_1),\ldots,\lambda_K/(1-\mu_K)$ so that

(1) $\displaystyle\sum_{n=1}^{K}\frac{\lambda_n}{1-\mu_n}=P_1$

(2) $\displaystyle\frac{1}{2}\sum_{n=1}^{K}\log\left(1+\frac{\lambda_n}{1-\mu_n}\right)$ is maximized, subject to (1).

We rewrite the following:

$$\begin{aligned}
&\frac{1}{2}\sum_{n=1}^{K}\log\left(1+\frac{\lambda_n}{1-\mu_n}\right) \\
&\quad=\frac{1}{2}\left\{K\sum_{n=1}^{K}\frac{1}{K}\log\left(1+\frac{\lambda_n}{1-\mu_n}\right)\Big/(K+P_1)+K\log(K+P_1)\right\}.
\end{aligned}$$

The second term is independent of the distribution of $\lambda_1/(1-\mu_1),\ldots,\lambda_K/(1-\mu_K)$. The first term is maximized if $(1+\lambda_n/(1-\mu_n))/(K+P_1)=1/K$ for $n=1,\ldots,K$ is a probability density when $\sum_{n=1}^{K}\lambda_n/(1-\mu_n)=P_1$. This gives $\lambda_n/(1-\mu_n)=P_1/K$ for $n=1,\ldots,K$. For fixed $K\leqslant M$ and $P_1\leqslant P$, we have that the maximum of $\frac{1}{2}\sum_{n=1}^{K}\log(1+(\lambda_n/(1-\mu_n)))$ is $(K/2)\log(1+(P_1/K))$. This maximum is easily seen to be maximized when $P_1=P$ and $K=M$. Hence we have obtained

$$\sup\{I(\mu_{12});\ \mu_{12}\in Q\}=(M/2)\log(1+(P/M)).$$

Moreover, the supremum is attained if, for example,

$$(S+V^*)(S^*+V)=\frac{2P}{M+2P}\sum_{n=1}^{M}e_n\otimes e_n$$

and

$$V^*V = \frac{P}{M+2P} \sum_{n=1}^{M} e_n \otimes e_n,$$

where $\{e_n\}$ are the eigenvectors of V^*V and $e_n \otimes e_n(x) = \langle x, e_n \rangle e_n$. Q.E.D.

When $\dim[H] = \infty$, the next theorem holds.

THEOREM 4. *Suppose that $M = \infty$ and $P > 0$, then $C(Q) = P/2$ and the supremum is not attained.*

Proof. As with Theorem 3, let $\{\lambda_n/(1-\mu_n)\}$ be, not necessarily finite, nonzero eigenvalues of $(I - V^*V)^{-1/2}\,(S + V^*)(S^* + V)(I - V^*V)^{-1/2}$. By Lemma 2,

$$I(\mu_{12}) = -\frac{1}{2}\sum_n \log\left(1 - \frac{\lambda_n}{1+\lambda_n-\mu_n}\right) = \frac{1}{2}\sum_n \log\left(1 + \frac{\lambda_n}{1-\mu_n}\right).$$

Constraint (c) implies that $\sum_n \lambda_n/(1-\mu_n) \leqslant P$. Then, for any $\mu_{12} \in Q$,

$$\frac{1}{2}\sum_n \log\left(1 + \frac{\lambda_n}{1-\mu_n}\right) \leqslant \frac{1}{2}\sum_n \frac{\lambda_n}{1-\mu_n} \leqslant \frac{P}{2}. \tag{*}$$

Suppose that $(I - V^*V)^{-1/2}\,(S + V^*)(S^* + V)(I - V^*V)^{-1/2}$ has K nonzero eigenvalues, then by the proof of Theorem 3, we know that the maximum of the average mutual information is $(K/2)\log(1 + (P/K))$. As K is arbitrary, we obtain $\lim_{K\to\infty}(K/2)\log(1 + (P/K)) = P/2$. Hence, all the inequalities in (*) must hold with the equalities. Whence $\sup\{I(\mu_{12});\ \mu_{12} \in Q\} = P/2$. It is easy to see that the supremum cannot be attained. Because if the supremum can be attained, then $\sum_n \log(1 + (\lambda_n/(1-\mu_n))) = \sum_n (\lambda_n/(1-\mu_n))$ and $\lambda_n/(1-\mu_n) = 0$ for all n. This implies $P = 0$ and induce the contradiction.

Q.E.D.

ACKNOWLEDGMENTS

The author should like to express his hearty thanks to Professor Hisaharu Umegaki of Tokyo Institute of Technology for his valuable suggestion and encouragement in the course of preparing this paper.

REFERENCES

1. R. B. ASH, "Information Theory," Interscience, New York, 1965.
2. C. R. BAKER, Mutual information for Gaussian processes, *SIAM J. Appl. Math.* **19** (1970), 451–458.

3. C. R. BAKER, On equivalence of probability measures, *Ann. Probab.* **1**(1973), 690–698.
4. C. R. BAKER, Joint measures and cross-covariance operators, *Trans. Amer. Math. Soc.* **186** (1973), 273–289.
5. C. R. BAKER, Zero-one laws for Gaussian measures on Banach spece, *Trans. Amer. Math. Soc.* **186** (1973), 290–308.
6. C. R. BAKER, Absolute continuity and applications to information theory, *in* "Probability in Banach Spaces," Lecture Notes in Mathematics, No. 526, Springer Pub., New York 1976.
7. C. R. BAKER, Capacity of the Gaussian channel without feedback, *Inform. and Control* **37** (1978), 70–89.
8. C. R. BAKER, Relations between mutual informations, strong equivalence, signal-to-ratio, and signal sample path properties, *in* "Transactions of the 8th Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes," Vol. A, pp. 85–92, 1978.
9. C. R. BAKER, Mutual information, strong equivalence, and signal sample path properties for Gaussian processes, *Inform. and Control* **41** (1979), 156–164.
10. R. L. DOBRUSHIN, General information of Shannon basic theorems on the theory of information, *Amer. Math. Soc. Transl.* **33** (2) (1963), 323–438.
11. R. G. DOUGLAS, On majorization, factorization, and range inclusion of operators in Hilbert space, *Proc. Amer. Math. Soc.* **17** (1966), 413–415.
12. I. M. GEL'FAND AND A. M. YAGLOM, Calculation of the amount of information about a random function contained in another such function, *Amer. Math. Soc. Transl.* **12** (12) (1959), 199–246.
13. J. HAJEK, On linear statistical problems in stochastic processes, *Czechoslovak Math. J.* **87** (1962), 404–444.
14. M. S. PINSKER, "Information and Information Stability for Random Variable and Processes," Holden–Day, San Francisco, 1964.
15. T. S. PITCHER, On the sample function of processes which can be added to a Gaussian process, *Ann. Math. Statist.* **34** (1963), 329–333.
16. C. R. RAO AND V. S. VARADARAJAN, Discrimination of Gaussian processes, *Sankhyā Ser. A* **25** (1963), 303–330.
17. R. SCHATTEN, "Norm Ideals of Completely Continuous Operators," Springer-Verlag, Berlin/New York, 1960.
18. A. V. SKOROHOD, "Integration in Hilbert Space," Springer-Verlag, Berlin/New York, 1974.
19. H. UMEGAKI, Absolute continuity of information channels, *J. Multivariate Anal.* **4** (1974), 382–400.